ARKIV FÖR FYSIK Band 17 nr 14

Communicated 2 December 1959 by H. Cramér and O. Klein

Macroscopic observables and generalized canonical ensembles

By Ulf Uhlhorn

1. Introduction

In the present note some aspects of the statistical description of macroscopic systems are discussed. What we are aiming at is a rational foundation, on a microscopic basis, of the theory of non equilibrium phenomena. However, this problem is only touched upon in the present work. For the characterization of macroscopic states non equilibrium thermodynamics uses observables which are closely related to the observables of equilibrium thermodynamics. For the complete characterization of macroscopic states in the sense of Ehrenfest's postulate of determination [Ehrenfest, 1911], that is the postulate that the macroscopic state of the system at a given time determines the macroscopic state at every later time, it is in general not sufficient to use observables of the thermodynamic type. It is assumed that it is possible to obtain a complete set of observables by adding to the observables of the thermodynamic type observables of other types. Ehrenfest's postulate refers to closed systems and for open systems it is obviously necessary to prescribe boundary conditions in order to make the characterization of the state complete. On a microscopic level the observables of the phenomenological theory are assumed to correspond to phase functions or to quantum mechanical operators. Obviously these functions and these operators must be of a very peculiar nature as they are responsible for the macroscopic properties of the system, in particular for its possible irreversible appearance.

The point of view taken in this note is that at least certain classes of macroscopic observables can be regarded as generators of generalized canonical distributions—the concept canonical distribution or ensemble being used in the sense of [Gibbs, 1902]—in a phase space, which may be of the classical type or of a more general nature. For this purpose it seemed desirable to investigate the physical and mathematical properties of such generalized canonical distributions in some detail. The content of this note is hence related both to the work of Grad on the statistical mechanics of systems with an arbitrary number of integrals of motion [Grad, 1952], and to the work of Münster on the transformation theory for the distributions of statistical mechanics [Münster, 1953]. It should be noticed, however, that the phase functions which we consider are not assumed to be even approximate integrals of motion. In fact the developments in this note are independent of dynamical considerations. The character of the macroscopic observables as functions of the time will be discussed in a forthcoming paper.

Asymptotic expansions of the probability distributions of macroscopic observables in terms of a parameter characterizing in some way the size of the system are very closely related to the classical limit theorems for sums of independent random variables. In particular an expansion given by Cramér [Cramér, 1928] appears to be useful for the statistical mechanics of macroscopic systems.

In the second paragraph of the note some basic phenomenological concepts are introduced. The general framework for the statistical description is presented in paragraph three. In paragraph four the macroscopic and microscopic descriptions are connected by the introduction of the generalized canonical distribution. The induced probability distributions for the macroscopic observables are discussed in paragraphs five, six and ten. Generalized thermodynamic relations are derived in paragraphs seven and eight. In paragraph nine stability conditions are derived from the generalized canonical distribution. Asymptotic expansions for the so called structure function $V(\cdot)$ are discussed in paragraph eleven. The structure function corresponds in the classical case to the phase volume between two energy surfaces. In an appendix the mathematical structure of the generalized canonical distribution is investigated.

2. Thermodynamic states

Let the thermodynamic state of a given system be characterized by the values of a finite number of extensive variables $(A^1, \ldots, A^n) \equiv A$:

$$A = \alpha \equiv (\alpha^1, \ldots, \alpha^n), \tag{1}$$

the thermodynamic coordinates of the system. It is assumed that every thermodynamic state (1) is a possible equilibrium state if the system is subjected to appropriate constraints. The characterization of thermodynamic states by the variables A thus corresponds to some prescribed class of constraints. If the constraints are changed within this class there will in general occur a change of the systems state. However, it is not assumed in the definition of thermodynamic equilibrium that for fixed constraints the system will ultimately reach an equilibrium state.

Illustrative examples of the extensive A-variables are energy, mass of substance, electric charge, of the total system and of subsystems.

The operational characterization of thermodynamic states by means of a prescribed class of constraints can be given a more precise meaning in the framework of statistical thermodynamics. From now on we shall consider the thermodynamics of the system as derived from a statistical model which will be described in the next paragraph.

3. Statistical thermodynamics

Let ζ denote the phase of the system which is macroscopically described by the variables A. We shall take the concept phase in a very wide sense. One may think of ζ as representing e.g. the number of particles of different species, their coordinates and conjugated momenta. One may also think of a quantum mechanical description of the system. In this case ζ will correspond to some representation of the linear operators in the state vector space which correspond to observables of the system. The detailed structure of the phase is, however, of minor importance for the following development. The essential property of the phase space $Z=\{\zeta\}$, considered as an abstract space, is that it has a measure $\mu(d\zeta)$ defined on it, corresponding to an integral

$$\int_Z x(\zeta)\,\mu(d\zeta), \tag{2}$$

where $x(\zeta)$ denotes an arbitrary function of the phase.[1] In general we have $\mu(Z)=+\infty$.

Any μ-continuous probability measure in Z has the form

$$\exp\left\{-\frac{1}{\varkappa}\sigma(\zeta)\right\}\mu(d\zeta), \tag{3}$$

where Boltzmann's constant $\varkappa$ has been introduced for convenience. The function $\sigma(\zeta)$, which is admitted to take the value $+\infty$, is subject only to the normalization condition

$$\int_Z \exp\left\{-\frac{1}{\varkappa}\sigma(\zeta)\right\}\mu(d\zeta)=1. \tag{4}$$

4. Correspondence postulate

The connection between the thermodynamic description and the phase space description is established by the following postulate.

Every macroscopic constraint of the specified class (paragragh 2) is characterized by the values of n real variables $(B_1, \ldots, B_n)\equiv B$:

$$B=\beta\equiv(\beta_1, \ldots, \beta_n). \tag{5}$$

---

[1] The precise formulation of these statements is: The phase space is a measure space

$$\{Z, \mathfrak{F}, \mu(\cdot)\}.$$

Here $\mathfrak{F}$ is a σ-field of measurable subsets of Z, containing Z and the empty set ϕ. The measure $\mu(\cdot)$ is a σ-additive, non negative set-function on $\mathfrak{F}$. In general $\mu(Z)=+\infty$. In this case $\mu(\cdot)$ is assumed to be σ-finite. In the future we shall avoid these technical terms which are explained in every text on measure theory, see e.g. [Halmos, 1950]. In particular when speaking about arbitrary functions of the phase we shall always mean real functions measurable with respect to $\mathfrak{F}$. It should be recognized that integration with respect to the measure $\mu(\cdot)$ will in general include ordinary integration with respect to the "Liouville measure" const. $d\,p\,d\,q$ and summations. In particular when a quantum mechanical description is used it is natural to consider the integral (2) as primarily defined as a linear functional on the space of functions of the phase and the measure $\mu(\cdot)$ as a derived quantity.

The domain of the B-variables, D_β is a convex subset of the n-dimensional real space R^n.[1] Every set of values of the B-variables determines a probability distribution $P_\beta(d\zeta)$ in Z.

To every extensive variable A^μ (paragraph 2) there corresponds a function of the phase $a^\mu(\zeta)$. Subject to the probability distribution $P_\beta(d\zeta)$, $a^\mu(\zeta)$ is a random variable. The values of this random variable are identified with the observed values of A^μ, given the constraint $B=\beta$. The thermodynamic state (paragraph 2) corresponding to the constraint $B=\beta$ is characterized by the expected values of the random variables $a^\mu(\zeta)$

$$\alpha^\mu = E_\beta a^\mu \equiv \int_Z a^\mu(\zeta)\, P_\beta(d\zeta), \qquad (\mu=1, \dots, n). \tag{6}$$

The probability measure $P_\beta(d\zeta)$ is of the form (3) with

$$\sigma_\beta(\zeta) \equiv \beta_\mu a^\mu(\zeta) - Y(\beta), {}^2 \tag{7}$$

that is $P_\beta(d\zeta)$ is a generalized canonical distribution.

The following conditions a., b., c. must be satisfied for every permissible β, that is for every $\beta \in D_\beta$:[3]

a. The function $\sigma_\beta(\zeta)$ defined by (7), must satisfy the normalization condition (4)

$$\exp\left\{-\frac{1}{\varkappa} Y(\beta)\right\} = \int_Z \exp\left\{-\frac{1}{\varkappa}\beta_\mu a^\mu(\zeta)\right\} \mu(d\zeta) < +\infty. \tag{8}$$

The normalization constant $Y(\beta)$ will be called the Massieu-Planck function [Planck, 1932], [Münster, 1959].

b. For arbitrary real constants $\lambda_1, \dots, \lambda_n$

$$\int_Z |\lambda_\mu a^\mu(\zeta)|^2 \exp\left\{-\frac{1}{\varkappa}\beta_\mu a^\mu(\zeta)\right\} \mu(d\zeta) < +\infty. \tag{9}$$

c. The random variables $a^\mu(\zeta)$ are linearly independent, that is, no relation of the form

$$\lambda_\mu a^\mu(\zeta) = \lambda_0, \tag{10}$$

where $\lambda_0, \lambda_1, \dots, \lambda_n$ are constants (independent of ζ), holds with probability one in Z unless $\lambda_0 = \lambda_1 = \dots = \lambda_n = 0$.

If the functions $a^\mu(\zeta)$ are finite almost everywhere in Z, that is if

$$|a^\mu(\zeta)| < \infty \qquad (\mu=1, \dots, n)$$

---

[1] This is equivalent to the statement: If $\beta 1 \in D_\beta$ and $\beta 2 \in D_\beta$, and if $0 \leqslant p \leqslant 1$, then $p\beta 1 + (1-p)\beta 2$, which is the point with coordinates $p\beta 1_\mu + (1-p)\beta 2_\mu$, is also contained in D_β; that is, with any two points, D_β contains the line segment joining these points.

[2] Summation convention is applied for pairs of equal upper and lower indices; the summation goes from 1 to n. If the sum is restricted, this is explicitely indicated, e.g. $\sum_{\mu=1}^{\mu=m} \alpha^\mu \beta_\mu$.

[3] See also appendix.

for every ζ with the possible exception of a ζ-set with μ-measure zero, and this will usually be the case in the applications, the condition c. can be replaced by the generally stronger statement that no relation of the form (10) can hold almost everywhere in Z. The latter statement is independent of β. For the proof denote by F the ζ-set for which (10) is true. Condition c. states

$$0 = P_\beta(F) \equiv \int_F \exp\left\{-\frac{1}{\varkappa}\sigma_\beta(\zeta)\right\}\mu(d\zeta),$$

and the finiteness assumption implies $|\sigma_\beta(\zeta)| < +\infty$ so that the non negative integrand does not vanish. But then we have necessarily

$$\mu(F) = 0,$$

which was to be proved. The meaning of c., is thus simply to exclude redundant variables.

According to the postulate a constraint of the specified class determines a thermodynamic state. The converse is also true. The statistical characterization of the constraints in terms of generalized canonical distributions will of course be possible only under particular experimental conditions and with a limited degree of accuracy. As the functions $a^\mu(\zeta)$ are not necessarily integrals of motion the probability measure $P_\beta(d\zeta)$ will in general not be a stationary measure for a closed system, that is a system described dynamically by a time independent Hamiltonian, invariant under the time reversal operation. The measure may be stationary for certain values of the parameters.

5. The probability distribution of the extensive variables

A possible value of the random variable $a(\zeta)$ is a point in the n-dimensional real space R^n. The phase measure $\mu(d\zeta)$ and the functions $a^\mu(\zeta)$ induce a measure in R^n by

$$V(da^1 \dots da^n) = \mu\{\zeta \,|\, a^\mu(\zeta) \in da^\mu,\ \mu = 1, \dots, n\}, \tag{11}$$

so that for an arbitrary function $f(a^1, \dots, a^n)$ on R^n

$$\int_{R^n} f(a^1, \dots, a^n)\, V(da^1 \dots da^n) = \int_Z f[a^1(\zeta), \dots, a^n(\zeta)]\,\mu(d\zeta). \tag{12}$$

Occasionally we shall assume that $V(da^1 \dots da^n)$, the structure function, is of the form

$$V(da^1 \dots da^n) = \exp\left\{\frac{1}{\varkappa}\varphi(a)\right\} da^1 \dots da^n, \tag{13}$$

$da^1 \dots da^n$ denoting the ordinary (Lebesgue) measure in R^n.

Given the probability measure (3), (7) we want to determine the probability distribution of the variables $a^1(\zeta), \dots, a^n(\zeta)$ in R^n. This distribution can be determined from the cumulant generating function [Cramér, 1937] defined by

$$C_\beta(iz) = \log E_\beta \exp\{iz_\mu a^\mu\} \equiv \log \int_Z \exp\{iz_\mu a^\mu(\zeta)\} P_\beta(d\zeta). \tag{14}$$

Here $z_1, \ldots, z_n$ are real parameters. It follows at once from the definition of $P_\beta(d\zeta)$ that

$$C_\beta(iz) = \frac{1}{\varkappa}\{Y(\beta) - Y(\beta - \varkappa iz)\} \tag{15}$$

if the definition of $Y(\beta)$ is extended to the complex plane. To obtain the probability distribution, let D be a bounded subset of R^n and define $e(D, z)$ by

$$e(D, z) = (2\pi)^{-n} \int_D \exp\{-iz_\mu x^\mu\} dx^1 \ldots dx^n. \tag{16}$$

Let $F_\beta(D)$ denote the probability of finding a value of $a(\zeta)$ in D, then [Cramér, 1937]

$$F_\beta(D) = \int_{R^n} e(D, z) \exp\{C_\beta(iz)\} dz_1 \ldots dz_n. \tag{17}$$

It is seen from (15) that the probability distribution $F_\beta(da^1 \ldots da^n)$ is completely determined by the Massieu-Planck function in the complex plane. On the other hand $F_\beta(da^1 \ldots da^n)$ can be expressed in terms of $V(da^1 \ldots da^n)$

$$F_\beta(da^1 \ldots da^n) = \exp\left\{-\frac{1}{\varkappa}[\beta_\mu a^\mu - Y(\beta)]\right\} V(da^1 \ldots da^n). \tag{18}$$

Accepting (13), we obtain from (17) and (18)

$$F_\beta(da^1 \ldots da^n) = \exp\left\{-\frac{1}{\varkappa}[\beta_\mu a^\mu - Y(\beta) - \varphi(a)]\right\} da^1 \ldots da^n, \tag{19}$$

$$\exp\left\{-\frac{1}{\varkappa}[\beta_\mu a^\mu - Y(\beta) - \varphi(a)]\right\} = (2\pi)^{-n} \int_{R^n} \exp\{C_\beta(iz) - iz_\mu a^\mu\} dz_1 \ldots dz_n \tag{20}$$

or

$$\exp\left\{\frac{1}{\varkappa}\varphi(a)\right\} = (2\pi)^{-n} \int_{R^n} \exp\left\{-\frac{1}{\varkappa}[Y(\beta - \varkappa iz) - (\beta_\mu - \varkappa iz_\mu) a^\mu]\right\} dz_1 \ldots dz_n. \tag{21}$$

It follows from the conditions a. and b. of paragraph 4. that for arbitrary constants $\lambda_1, \ldots, \lambda_n$, $E_\beta |\lambda_\mu a^\mu|^2 < \infty$, and by Schwarz' inequality $E_\beta |\lambda_\mu a^\mu| < \infty$. Thus the mean

$$E_\beta \lambda_\mu a^\mu = \lambda_\mu E_\beta a^\mu = \lambda_\mu \alpha^\mu, \tag{22}$$

and the variance

$$E_\beta |\lambda_\mu a^\mu - E_\beta \lambda_\nu a^\nu|^2 = \varkappa f^{\mu\nu} \lambda_\mu \lambda_\nu, \tag{23}$$

exist for arbitrary real $\lambda_1, \ldots, \lambda_n$. The quantity $\varkappa f^{\mu\nu} = \varkappa f^{\nu\mu}$ is called the covariance of $a^{\mu}(\zeta)$ and $a^{\nu}(\zeta)$, $\mathrm{Cov}_{\beta}(a^{\mu}, a^{\nu})$. Being the expected value of a non negative quantity (23) the quadratic form corresponding to $\varkappa f^{\mu\nu}$ is non negative

$$\varkappa f^{\mu\nu} \lambda_{\mu} \lambda_{\nu} \geqslant 0. \tag{24}$$

The condition c. of paragraph 4. is equivalent to the statement that the quadratic form (24) is positive definite, that is

$$\varkappa f^{\mu\nu} \lambda_{\mu} \lambda_{\nu} = 0 \qquad \text{implies } \lambda_1 = \cdots = \lambda_n = 0. \tag{25}$$

If not, there would exist a linear combination of the $a^{\mu}(\zeta)$ with vanishing variance (23), but that is equivalent to the fact that with probability one this linear combination is equal to its mean value, contrary to the hypothesis c.

The existence of the mean values and covariances being established it follows from wellknown, easily verifiable properties of the cumulant generating function [Cramér, 1937] that

$$\alpha^{\mu} = E_{\beta} a^{\mu} = \frac{\partial C_{\beta}}{\partial i z_{\mu}}(0) = \frac{\partial Y(\beta)}{\partial \beta_{\mu}}, \tag{26}$$

$$\varkappa f^{\mu\nu} = \mathrm{Cov}_{\beta}(a^{\mu}, a^{\nu}) = \frac{\partial^2 C_{\beta}}{\partial i z_{\mu} \partial i z_{\nu}}(0) = -\frac{\partial^2 Y(\beta)}{\partial \beta_{\mu} \partial \beta_{\nu}}. \tag{27}$$

6. Approximation to the probability distribution of the extensive variables obtained from the central limit theorem

For the further discussion of the probability distribution of the random variables $a^1(\zeta), \ldots, a^n(\zeta)$ it is convenient to introduce the variable

$$x(\zeta) = \frac{x_{\lambda}[a^{\lambda}(\zeta) - \alpha^{\lambda}]}{\sqrt{\varkappa f^{\mu\nu} x_{\mu} x_{\nu}}}, \tag{28}$$

where x_{λ} are arbitrary constants, not all equal to zero. The variable $x(\zeta)$ is standardized

$$E_{\beta} x = 0, \qquad E_{\beta} x^2 = 1. \tag{29}$$

The cumulant generating function corresponding to $x(\zeta)$ is

$$\theta(ik) = \log E_{\beta} \exp ikx = -ik \frac{x_{\lambda} \dfrac{\partial Y}{\partial \beta_{\lambda}}}{d} + \frac{1}{\varkappa}\left\{Y(\beta) - Y\left(\beta - \varkappa \frac{ikx}{d}\right)\right\} \tag{30}$$

where

$$d = \sqrt{-\varkappa \frac{\partial^2 Y}{\partial \beta_{\mu} \partial \beta_{\nu}} x_{\mu} x_{\nu}} \tag{31}$$

and k is a real parameter.

If they exist, the cumulants of the distribution are defined as the derivatives of $\theta(ik)$ taken at $ik = 0$. The cumulants can be expressed as polynomials in the moments of the distribution. We have for the first few, taking into account that x is standardized [Cramér, 1945]

$$\theta^{(1)}(0) = 0 \tag{32}$$

$$\theta^{(2)}(0) = 1 \tag{33}$$

$$\theta^{(3)}(0) = E_\beta x^3 \tag{34}$$

$$\theta^{(4)}(0) = E_\beta x^4 - 3. \tag{35}$$

And combining this with (30) we find

$$E_\beta x^3 = -\frac{\varkappa^2}{d^3} x_\lambda x_\mu x_\nu \frac{\partial^3 Y(\beta)}{\partial\beta_\lambda \partial\beta_\mu \partial\beta_\nu}, \tag{36}$$

$$E_\beta x^4 = 3 + \frac{\varkappa^3}{d^4} x_\iota x_\lambda x_\mu x_\nu \frac{\partial^4 Y(\beta)}{\partial\beta_\iota \partial\beta_\lambda \partial\beta_\mu \partial\beta_\nu}. \tag{37}$$

The cumulant generating function corresponding to the standardized Gauszian distribution

$$d\Phi(y) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2} y^2\right\} dy \tag{38}$$

is

$$\log \int d\Phi(y) \exp\{iky\} = \frac{1}{2}(ik)^2. \tag{39}$$

The second cumulant of the Gauszian distribution is equal to one, and all the others are zero. Thus the cumulants of an arbitrary standardized distribution describe in a systematic way its deviation from the Gauszian form.

It is obviously not possible to procede further without making specific assumptions about the form of the Massieu-Planck function $Y(\beta)$ which determines the probability distribution of $x(\zeta)$. From the point of view of statistical thermodynamics this amounts to a specification of the phase space Z and the functions $a^\mu(\zeta)$. However, except for the case when the $a(\zeta)$ happen to be sums of many stochastically independent and equally or almost equally distributed random variables we are not able to determine the form of $Y(\beta)$. On the other hand, adopting an "Einsteinian point of view" [Einstein, 1910], and considering the Massieu-Planck function $Y(\beta)$ as an empirically determined thermodynamic quantity, we would expect $Y(\beta)$, as an extensive variable, to be proportional to the size of the system. So, assuming that all the derivatives of the function with respect to the intensive parameters are of the same order of magnitude as the function itself, we would expect the r:th cumulant to be of the order $\sim \left[\frac{|Y(\beta)|}{\varkappa}\right]^{1-r/2}$. For macroscopic systems the higher order cumulants would then be small, and the Gauszian distribution would be a good approximation for the random variable $x(\zeta)$. Deviations from the Gauszian form could be determined from (36), (37).

A more precise formulation of the latter considerations is that $Y(\beta)$ can be written in the form $N Y_1(\beta)$ corresponding to a possible decomposition of the system in a great number N of, still macroscopic, subsystems which have identical Massieu-Planck functions. The assumption that the subsystems are macroscopic is necessary to assure that the Massieu-Planck function for the entire system is approximately equal to the sum of the corresponding functions for the subsystems. The assumption that the subsystem have identical Massieu-Planck functions can be replaced by various somewhat weaker homogenity conditions which we shall not state explicitly. They correspond to different formulations of the central limit theorem for sums of independent not equally distributed random variables [Gnedenko and Kolmogorov, 1954]. It is the first assumption, corresponding to stochastic independence of the subsystems, which is most difficult to justify. What one would actually need is a central limit theorem for sums of dependent variables and such a theorem has not been proved at present.[1] Taking the assumptions for granted we find that when the number of subsystems tends to infinity, the cumulant generating function for the distribution of x will for fixed k converge to the form (39). It follows [Cramér, 1945] that the distribution function for $x(\zeta)$, i.e. the integral of the probability density, will converge to the Gauszian form for every value of its argument. The purpose of this vague and heuristic discussion was only to indicate a way in which the problem of finding the asymptotic distribution of x for large systems could possibly be traced back to the classical form of the central limit theorem for sums of independent variables. It is however possible that an actual derivation of the asymptotic distribution must procede along completely heterodox lines. Before discussing the Gauszian approximation to the probability distribution of the extensive variables we shall derive some generalized thermodynamic relations.

7. Thermodynamic relations

According to the postulate in paragraph 4 a constraint of the specified class is characterized by a set of values of the intensive variables $B=\beta$, $\beta \in D_\beta$, and determines uniquely a thermodynamic state with coordinates $\alpha^\mu = E_\beta a^\mu$. Conversely the thermodynamic state determines the constraint uniquely. The corresponding values β of the parameters B are obtained by inverting the relations (26)

$$\alpha^\mu = \frac{\partial Y}{\partial \beta^\mu}(\beta) \qquad (\mu = 1, \ldots, n).^{2} \tag{40}$$

---

[1] There are versions of the central limit theorem which consider sums of stochastically dependent random variables. But the dependence is of a very weak kind so that the variables in the sum are asymptotically independent. A theorem of this type holds e.g. for the energy of one independent component of an assembly of non interacting systems when the probability distribution is microcanonical and the number of systems becomes large. A particular case of this theorem was proved already by Maxwell [Maxwell, 1878].

[2] As is wellknown the non vanishing of the Jacobian $\dfrac{\partial(\alpha^1, \ldots, \alpha^n)}{\partial(\beta_1, \ldots, \beta_n)}$ is a sufficient condition for the existense of an inverse transformation. However, the condition is not sufficient to assure the uniqueness of the inverse transformation (simple example $y=x^2$). The latter fact is forgotten in some texts on thermodynamics which at all discuss the existence of the inverse. The essential property used in the following proof is the convexity (24), (25), (27) of the thermodynamical potentials, that is the convex form of the thermodynamic surfaces.

This can be done uniquely for

$$\frac{\partial Y}{\partial \beta_\mu}(\beta 1) = \frac{\partial Y}{\partial \beta_\mu}(\beta 2) \qquad (\mu = 1, \ldots, n)\, \beta 1,\ \beta 2 \in D_\beta \text{ implies } \beta 1 = \beta 2. \tag{41}$$

To see this let $\beta 1 \neq \beta 2$ be two points in D_β and consider the joining segment, which is contained in D_β as D_β is convex: $\beta(p) = p\beta 1 + (1-p)\beta 2$, $0 \leqslant p \leqslant 1$. The function $f(p) = Y[\beta(p)]$ has the derivatives

$$f'(p) = \frac{\partial Y}{\partial \beta_\mu} \beta(p)'_\mu, \quad \text{and} \quad f''(p) = \frac{\partial^2 Y}{\partial \beta_\mu \partial \beta_\nu} \beta(p)'_\mu \beta(p)'_\nu$$

with $\beta(p)' = \beta 1 - \beta 2$. According to (24), (25), (27) the second derivative is always negative so that $f'(0) > f'(1)$ or

$$\left\{\frac{\partial Y}{\partial \beta_\mu}(\beta 1) - \frac{\partial Y}{\partial \beta_\mu}(\beta 2)\right\}(\beta 1_\mu - \beta 2_\mu) > 0,$$

and it follows that

$$\frac{\partial Y}{\partial \beta_\mu}(\beta 1) \neq \frac{\partial Y}{\partial \beta_\mu}(\beta 2)$$

for at least one value of μ. The statement (41) follows.

Hence the Legendre transformation [e.g. Goursat, 1917]

$$\begin{aligned} \alpha_\mu &= \frac{\partial Y}{\partial \beta_\mu}(\beta) \\ S &= \frac{\partial Y}{\partial \beta_\mu} \beta_\mu - Y \end{aligned} \tag{42}$$

is well defined and reversible

$$\begin{aligned} \beta_\mu &= \frac{\partial S}{\partial \alpha^\mu}(\alpha) \\ Y &= \frac{\partial S}{\partial \alpha^\mu} \alpha^\mu - S. \end{aligned} \tag{43}$$

The system (42), (43) illustrates the thermodynamic signification of the Massieu-Planck function as $S(\alpha)$ can be identified with the entropy function of the system, being equal to the expected value of the probability exponent

$$S(\alpha) = E_\beta\, \sigma_\beta = E_\beta[\beta_\mu a^\mu - Y] = \beta_\mu \alpha^\mu - Y. \tag{44}$$

The following relation between the entropy and the Massieu-Planck function is a general property of the Legendre transformation

$$\frac{\partial^2 S}{\partial \alpha^\lambda \partial \alpha^\mu} \frac{\partial^2 Y}{\partial \beta_\mu \partial \beta_\nu} = \delta^\nu_\lambda. \tag{45}$$

The relation is an immediate consequence of the fact that

$$\frac{\partial S}{\partial \alpha^{\lambda}} = \beta_{\lambda}(\alpha) \quad \text{and} \quad \frac{\partial Y}{\partial \beta_{\mu}} = \alpha^{\mu}(\beta).$$

The values of the extensive variables A or of the intensive variables B are equivalent characterizations of a thermodynamic state and of the corresponding probability measure in the phase space.

It follows exactly as above, as a consequence of the condition c. in paragraph 4., that it is possible to represent a thermodynamic state by the values

$$(\beta_1, \ldots, \beta_m, \alpha^{m+1}, \ldots, \alpha^n) \qquad (0 \leqslant m \leqslant n). \tag{46}$$

The corresponding Legendre transformation is

$$\begin{aligned} \beta_{\mu} &= \frac{\partial S}{\partial \alpha^{\mu}} \qquad (\mu = 1, \ldots, m) \\ Y_m &= \sum_{\mu=1}^{\mu=m} \frac{\partial S}{\partial \alpha^{\mu}} \alpha^{\mu} - S. \end{aligned} \tag{47}$$

and the inverse transformation

$$\begin{aligned} \alpha^{\mu} &= \frac{\partial Y_m}{\partial \beta_{\mu}} \qquad (\mu = 1, \ldots, m) \\ S &= \sum_{\mu=1}^{\mu=m} \frac{\partial Y_m}{\partial \beta_{\mu}} \beta_{\mu} - Y_m. \end{aligned} \tag{48}$$

The notation implies $Y_0 = -S$ and $Y_n = Y$. From (47) and (48) it follows that

$$d\,Y_m = \sum_{\mu=1}^{\mu=m} \alpha^{\mu}\, d\beta_{\mu} - \sum_{\mu=m+1}^{\mu=n} \beta_{\mu}\, d\alpha^{\mu}, \tag{49}$$

and that for $0 \leqslant l < m \leqslant n$

$$\{-Y_l, \alpha^{l+1}, \ldots, \alpha^m\} \quad \text{and} \quad \{Y_m, \beta_{l+1}, \ldots, \beta_m\} \tag{50}$$

are connected by the Legendre transformations of $m - l + 1$ variables

$$\beta_{\mu} = \frac{\partial(-Y_l)}{\partial \alpha^{\mu}} \ (\mu = l+1, \ldots, m), \ Y_m = \sum_{\mu=l+1}^{\mu=m} \frac{\partial(-Y_l)}{\partial \alpha^{\mu}} \alpha^{\mu} - (-Y_l), \tag{51}$$

$$\alpha^{\mu} = \frac{\partial Y_m}{\partial \beta_{\mu}} \ (\mu = l+1, \ldots, m), \ (-Y_l) = \sum_{\mu=l+1}^{\mu=m} \frac{\partial Y_m}{\partial \beta_{\mu}} \beta_{\mu} - Y_m. \tag{52}$$

It follows from the general property (45) of the Legendre transformation

$$\sum_{\mu=l+1}^{\mu=m} \frac{\partial^2(-Y_l)}{\partial \alpha^{\lambda} \partial \alpha^{\mu}} \frac{\partial^2 Y_m}{\partial \beta_{\mu} \partial \beta_{\nu}} = \delta_{\lambda}^{\nu} \qquad (\lambda, \nu = l+1, \ldots, m). \tag{53}$$

8. Linear transformations of thermodynamic coordinates

Thermodynamic constraints are expressed either as prescriptions for the values of the intensive B-variables or as linear conditions on the extensive A-variables, which obey conservation laws. A linear condition may have a particularly simple form in some coordinate system. We shall therefore consider non singular linear transformations of the extensive variables

$$\left.\begin{aligned} \tilde{\alpha}^{\mu} &= A^{\mu}_{\nu}\,(\alpha^{\nu} - \alpha^{\nu}_{0}) \\ \alpha^{\mu} &= \alpha^{\mu}_{0} + B^{\mu}_{\nu}\,\tilde{\alpha}^{\nu} \end{aligned}\right\}\; A^{\mu}_{\lambda}\,B^{\lambda}_{\nu} = B^{\mu}_{\lambda}\,A^{\lambda}_{\nu} = \delta^{\mu}_{\nu}. \tag{54}$$

The transformed entropy function $\tilde{S}\,(\tilde{\alpha})$ is given by

$$\tilde{S}\,(\tilde{\alpha}) = S\,(\alpha_0 + B\,\tilde{\alpha}), \tag{55}$$

and the corresponding intensive variables $\tilde{\beta}_{\mu}$ by

$$\tilde{\beta}_{\mu} = \frac{\partial\,\tilde{S}\,(\tilde{\alpha})}{\partial\,\tilde{\alpha}^{\mu}} = \beta_{\nu}\,B^{\nu}_{\mu}. \tag{56}$$

The Massieu-Planck function $\tilde{Y}\,(\tilde{\beta})$ corresponding to (55) and (56) is

$$\tilde{Y}\,(\tilde{\beta}) = \tilde{S}\,(\tilde{\alpha}) - \tilde{\alpha}^{\mu}\,\tilde{\beta}_{\mu} = S\,(\alpha) - A^{\mu}_{\nu}\,(\alpha^{\nu} - \alpha^{\nu}_{0})\,\beta_{\lambda}\,B^{\lambda}_{\mu} = Y\,(\beta) + \alpha^{\nu}_{0}\,\beta_{\nu}. \tag{57}$$

The corresponding transformation of the phasefunctions $a_{\mu}\,(\zeta)$ is

$$\tilde{a}^{\mu}\,(\zeta) = A^{\mu}_{\nu}\,[a^{\nu}\,(\zeta) - \alpha^{\nu}_{0}]. \tag{58}$$

It follows directly from (57) and (58) that

$$\sigma\,(\zeta) = Y\,(\beta) + \beta_{\mu}\,a^{\mu}\,(\zeta) = \tilde{Y}\,(\tilde{\beta}) + \tilde{\beta}_{\mu}\,\tilde{a}^{\mu}\,(\zeta), \tag{59}$$

which expresses the invariance of the probability measure under the combined transformation (58), (56), (57). The domain of permissible β-values is transformed by (56) into a domain $\tilde{D}_{\beta}$ which is obviously also convex.

It is clear that there is no ambiguity in expressing linear conditions on the A-variables in the form $\alpha^{m+1} = \cdots = \alpha^{n} = 0$.

9. Relaxation of constraints

A constraint of the prescribed class (paragraph 2) is characterized by the values of n parameters, either of the A-variables, or of the B-variables, or of a mixture of the A- and the B-variables. A different class of constraints is obtained if only a smaller number of parameters are specified, e.g. $B_1, \ldots, B_l$ and $A^{m+1}, \ldots, A^{n}$ with $0 \leqslant l < m \leqslant n$, and the variables $A^{l+1}, \ldots, A^{m}$ are left undetermined. A priori we do not know whether a constraint of this type determines a thermodynamic state or not. Having postulated a one to one correspondence between thermodynamic states and probability measures in the phase space we have to find out whether a probability exponent of the type

$$\sigma(\zeta) = c + \sum_{\mu=1}^{\mu=l} \beta_\mu a^\mu(\zeta) + \sum_{\mu=m+1}^{\mu=n} \beta_\mu a^\mu(\zeta) \tag{60}$$

can determine a probability measure in Z. Obviously (60) defines a probability measure if and only if

$$(\beta_1, \ldots, \beta_l, 0, \ldots, 0, \beta_{m+1}, \ldots, \beta_n) \in D_\beta, \tag{61}$$

and then

$$c = Y(\beta_1, \ldots, \beta_l, 0, \ldots, 0, \beta_{m+1}, \ldots, \beta_n). \tag{62}$$

Let the constraint be given in the form

$$B_1 = \beta_1, \ldots, B_l = \beta_l, \; A^{m+1} = \alpha^{m+1}, \ldots, A^n = \alpha^n. \tag{63}$$

The condition (61) is then equivalent to

$$\frac{\partial}{\partial \alpha^\mu} Y_l(\beta_1, \ldots, \beta_l, \alpha^{l+1}, \ldots, \alpha^m, \alpha^{m+1}, \ldots, \alpha^n) = \beta_\mu = 0 \quad (\mu = l+1, \ldots, m). \tag{64}$$

But (64) is the condition for Y_l to be stationary considered as a function of $\alpha^{l+1}, \ldots, \alpha^m$, the other variables being determined by the constraint (63). The stationary value is a minimum as follows directly from the fact that the second order differential is positive definite, this being a consequence of the assumption c. in paragraph 4. in the form (24), (25) and the relation (53) with $m = n$.

Thus we can state: The constraint (63) determines a thermodynamic state if and only if Y_l has a minimum when considered as a function of $\alpha^{l+1}, \ldots, \alpha^m$ under the given constraint. The minimum value is attained in this thermodynamic state. This is of course nothing but the thermodynamic stability condition, derived here from the postulate of paragraph 4. If in particular, only extensive variables are specified $l = 0$ and as $Y_0 = -S$, the condition for the existence of a thermodynamic state, is that S has a maximum as a function of the "free" variables.

10. Marginal distributions of the extensive variables. The Gauszian approximation

In this paragraph we shall conclude the discussion in paragraph 5. and paragraph 6. of the distribution of the extensive variables $a^1(\zeta), \ldots, a^n(\zeta)$. As we saw the n-dimensional probability distribution of these variables was obtained from the Massieu-Planck function by a Fourier transformation involving the cumulant generating function or equivalently the Massieu-Planck function (17), (15). If we are interested only in the distribution of some of the $a^\mu(\zeta)$:s say $a^{l+1}(\zeta), \ldots, a^m(\zeta)$ the other $a^\mu(\zeta)$:s being left unspecified, we put $z_1 = \cdots = z_l = z_{m+1} = \cdots = z_n = 0$ in (14) and (15) to obtain the cumulant generating function of the corresponding marginal distribution. If $z_{m+1} = \cdots = z_n = 0$ we may replace Y in (15) by Y_m and this function determines the $(m-l)$-dimensional probability distribution of $a^{l+1}(\zeta), \ldots, a^m(\zeta)$. In particular mean values and covariances are given by

$$E_\beta a^\mu = \frac{\partial Y_m}{\partial \beta_\mu} \qquad (\mu = l+1, \ldots, m), \tag{65}$$

$$\mathrm{Cov}_\beta (a^\lambda, a^\mu) = -\frac{\partial^2 Y_m}{\partial \beta_\lambda \partial \beta_\mu} \qquad (\lambda, \mu = l+1, \ldots, m). \tag{66}$$

It follows from (53) that

$$\sum_{\mu=l+1}^{\mu=m} \frac{\partial^2 Y_l}{\partial \alpha^\lambda \partial \alpha^\mu} \mathrm{Cov}_\beta (a^\mu, a^\nu) = \varkappa \, \delta_\lambda^\nu \, (\lambda, \nu = l+1, \ldots, m). \tag{67}$$

that is, the covariance is inverted by $\left\{\frac{1}{\varkappa} \frac{\partial^2 Y_l}{\partial \alpha^\lambda \partial \alpha^\mu} ; \lambda, \mu = l+1, \ldots, m\right\}$.

Let us return for a moment to the standardized random variable $x(\zeta)$ introduced in paragraph 6, (28). Its distribution is determined from the cumulant generating function (30). If an homogenity assumption of the form discussed in paragraph 6 is accepted, the cumulant generating function can be expanded formally in a Taylor series in powers of the inverse square root of the "size" N of the system. The zero order term corresponds to the standardized Gauszian distribution function

$$\Phi(x) = \int_{-\infty}^{x} \varphi(y)\, dy, \quad \varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\tfrac{1}{2} x^2\right). \tag{68}$$

To the Taylor expansion of the cumulant generating function there will then under fairly general conditions correspond an expansion of the frequency function $f(x)$ for $x(\zeta)$, the first two terms of which are

$$f(x) = \varphi(x) - \frac{\gamma_1}{3!} \varphi^{(3)}(x) + \frac{\gamma_2}{4!} \varphi^{(4)}(x) + \frac{10\gamma_1^2}{6!} \varphi^{(6)}(x) + \cdots \tag{69}$$

and which, as Cramér has shown [Cramér, 1928], really gives an asymptotic expansion in powers of $N^{-\frac{1}{2}}$, with a remainder term of the same order as the first term neglected. The parameters γ_1 and γ_2 which are given by

$$\gamma_1 = E_\beta x^3 = \theta^{(3)}(0)$$

$$\gamma_2 = E_\beta x^4 - 3 = \theta^{(4)}(0),$$

are usually called coefficient of skewness and coefficient of excess, respectively. It is interesting to notice that the expansion (69) can also be considered as an expansion in orthogonal Hermite polynomials.

On the same assumptions a corresponding expansion may be obtained for the frequency function for the variables $a^{l+1}(\zeta), \dots, a^m(\zeta)$. The zero order term is the Gauszian frequency function determined by the mean values (65) and the covariances (66). As a result of (67) it has the form

$$(2\pi)^{-\frac{m-l}{2}} \operatorname{Det}\left\{\frac{1}{\varkappa}\frac{\partial^2 Y_l}{\partial\alpha^\lambda\partial\alpha^\mu};\ \lambda,\mu=l+1,\dots,m\right\}^{\frac{1}{2}} \times$$
$$\times \exp\left\{-\frac{1}{2\varkappa}\sum_{\lambda,\mu=l+1}^{\lambda,\mu=m}\sum \frac{\partial^2 Y_l}{\partial\alpha^\lambda\partial\alpha^\mu}(a^\lambda-\alpha^\lambda)(a^\mu-\alpha^\mu)\right\}. \tag{71}$$

In particular for $l=0$ we have $Y_0=-S$ and

$$(2\pi)^{-n/2} \operatorname{Det}\left\{-\frac{1}{\varkappa}\frac{\partial^2 S}{\partial\alpha^\lambda\partial\alpha^\mu};\ \lambda,\mu=1,\dots,m\right\}^{\frac{1}{2}} \times$$
$$\times \exp\left\{+\frac{1}{2\varkappa}\sum_{\lambda,\mu=1}^{\lambda,\mu=m}\sum \frac{\partial^2 S}{\partial\alpha^\lambda\partial\alpha^\mu}(a^\lambda-\alpha^\lambda)(a^\mu-\alpha^\mu)\right\}. \tag{72}$$

If the l intensive variables $B_1, \dots, B_l$ and the $n-m$ extensive variables $A^{m+1}, \dots, A^n$ are fixed, the remaining free extensive variables $A^{l+1}, \dots, A^m$ are distributed around their mean values $\alpha^{l+1}, \dots, \alpha^m$ under the given constraint, which is assumed to determine a thermodynamic state, according to the probability distribution

$$C \exp\left\{-\frac{1}{\varkappa}\Delta Y_l(\Delta a)\right\} d(\Delta a^{l+1}) \dots d(\Delta a^m). \tag{73}$$

Here C is a normalization constant, i.e. independent of $a^{l+1}, \dots, a^m$, and $\Delta Y_l(\Delta a)$ is formally equal to the first two terms of the Taylor expansion of

$$Y_l(\beta_1, \dots, \beta_l, a^{l+1}, \dots, a^m, \alpha^{m+1}, \dots, \alpha^n) - Y_l(\beta_1, \dots, \beta_l, \alpha^{l+1}, \dots, \alpha^m, \alpha^{m+1}, \dots, \alpha^n)$$

in powers of the deviations from the means of the free variables $A^{l+1}, \dots, A^m$. As was demonstrated in paragraph 9, Y_l is the appropriate thermodynamic potential for the given type of constraint, and has a minimum in the corresponding thermodynamic state. All derivatives of the first order vanish and thus (73) is seen to be only a particular case of (71).

Corresponding to (72), $l=0$, $Y_0=-S$, we obtain from (73) a version of Boltzmann's principle: "Entropie = log (Wahrscheinlichkeit)". Here "Entropie" is to be interpreted as the first non vanishing term in the Taylor expansion of ΔS.

Objection could be raised against the way in which we have taken account of the constraint $A^{m+1}=\alpha^{m+1}, \dots, A^n=\alpha^n$ for the determination of the distribution (73). We have simply used the marginal distribution of the $A^{l+1}, \dots, A^m$ with parameters $\beta_1, \dots, \beta_n$ or equivalently $\beta_1, \dots, \beta_l$, $\alpha^{l+1}, \dots, \alpha^n$. In some situations it would be more natural to consider not the marginal distribution but the conditional probability distribution of the $A^{l+1}, \dots, A^m$. The condition should be of the form $A^{m+1} = \alpha^{m+1}, \dots, A^n = \alpha^n$. However, due to the fact that the distribution of $A^1, \dots, A^n$ is Gauszian we obtain in this way the same distribution for $A^{l+1}, \dots, A^m$ (71).

11. Asymptotic expansion of φ (*a*)

We shall assume that we are dealing with a system of such a nature that an expansion of the Massieu-Planck function in powers of the inverse square root of a dimensionless size parameter N is meaningful. We shall further assume that the measure $V(d a^1 \dots d a^n)$ on R^n has a density (13) $\exp\left\{\frac{1}{\varkappa}\varphi(a)\right\}$.

Given the constraint $B = \beta$, the probability density for the variables $A^1, \dots, A^n$ is, without approximation, given by

$$\exp\left\{\frac{1}{\varkappa}[\varphi(a) + Y(\beta) - \beta_\mu a^\mu]\right\} \equiv \exp\left\{\frac{1}{\varkappa}\left[\varphi(a) - S(\alpha) - \frac{\partial S}{\partial \alpha^\mu}(a^\mu - \alpha^\mu)\right]\right\}. \tag{75}$$

Corresponding to the asymptotic expansion of the probability density we may write (75) in the form

$$\exp\left\{\frac{1}{\varkappa}\frac{1}{2}\frac{\partial^2 S}{\partial \alpha^\mu \partial \alpha^\nu}(a^\mu - \alpha^\mu)(a^\nu - \alpha^\nu)\right\} C(\alpha)\{1 + k(a, \alpha)\}, \tag{76}$$

where

$$C(\alpha) = (2\pi)^{-n/2} \operatorname{Det}\left\{\frac{-\partial^2 S}{\partial \alpha^\mu \partial \alpha^\nu}\right\}^{\frac{1}{2}} \tag{77}$$

and the function $k(a, \alpha)$ is a power series in $N^{-\frac{1}{2}}$. Combining (75) and (76) we obtain the following expression for $\varphi(a)$

$$\begin{aligned}\varphi(a) = S(\alpha) + \frac{\partial S}{\partial \alpha^\mu}(a^\mu - \alpha^\mu) + \frac{1}{2}\frac{\partial^2 S}{\partial \alpha^\mu \partial \alpha^\nu}(a^\mu - \alpha^\mu)(a^\nu - \alpha^\nu) + \\ + \varkappa \log C(\alpha) + \varkappa \log\{1 + k(a, \alpha)\}.\end{aligned} \tag{78}$$

We recall that the terms in $k(a, \alpha)$ are n-dimensional Hermite polynomials with coefficients of definite orders in $N^{-\frac{1}{2}}$. The first term of the expansion of (78) in powers of $N^{-\frac{1}{2}}$ corresponding to $k(a, \alpha) = 0$, is

$$\varphi_0(a, \alpha) = S(\alpha) + \frac{\partial S}{\partial \alpha^\mu}(a^\mu - \alpha^\mu) + \frac{1}{2}\frac{\partial^2 S}{\partial \alpha^\mu \partial \alpha^\nu}(a^\mu - \alpha^\mu)(a^\nu - \alpha^\nu) + \varkappa \log C(\alpha). \tag{79}$$

We can replace $\varphi(a)$ by $\varphi_0(a, \alpha)$ with an error of the order $N^{-\frac{1}{2}}$ which is of course not uniform in a and α. It might be remarked that the usual derivation of this result in the case $n = 1$, $A^1 =$ the energy, e.g. [Gibbs, 1902] is purely formal and can give no estimate of the error.

The assumption that the exact form of $\varphi(a)$ is

$$\varphi(a) = \varphi_0 + \varphi_\mu a^\mu + \tfrac{1}{2}\varphi_{\mu\nu} a^\mu a^\nu \tag{80}$$

with constant coefficients $\varphi_0, \varphi_\mu, \varphi_{\mu\nu}$, is strictly incompatible with the basic assumption that

$$\int_{R^n} \exp\left\{\frac{1}{\varkappa}\varphi(a)\right\} d a^1 \ldots d a^n = +\infty. \tag{81}$$

Thus there is no system for which (80) is exactly valid. It has meaning only as an asymptotic result when, with α^μ given by $\varphi_{\mu\nu}\alpha^\nu = -\varphi_\mu$,

$$(2\pi)^{\frac{n}{2}} \operatorname{Det}\left\{\frac{1}{\varkappa}\varphi_{\mu\nu}\right\}^{\frac{1}{2}} \exp\left\{\frac{1}{\varkappa}\left[\varphi_0 - \frac{1}{2}\alpha^\mu \varphi_\mu\right]\right\} \to +\infty. \tag{82}$$

APPENDIX

The mathematical structure of the generalized canonical distribution

In this appendix the generalized canonical distribution of paragraph 4. is discussed in a less heuristic manner. A connection between the possible values of the phase functions and the possible mean values of the generalized canonical distribution, which generalizes results of Khinchin and Grad, is established. We shall take Boltzmann's constant $\varkappa = 1$.

To every extensive variable A^μ (paragraph 2), there corresponds a real, measurable function $a^\mu(\zeta)$ of the phase. We denote by

$$a(\zeta) \equiv \{a^1(\zeta), \ldots, a^n(\zeta)\} \tag{A 1}$$

the corresponding measurable function from Z into the n-dimensional real space R^n.

We shall assume that the functions $a^\mu(\zeta)$ are linearly independent. That is, if for some real constants $\lambda_0, \lambda_1, \ldots, \lambda_n$

$$\lambda_0 + \lambda_\mu a^\mu(\zeta) = 0 \tag{A 2}$$

holds for all points of the phase space Z with the possible exception of a set with μ-measure zero, it follows that

$$\lambda_0 = \lambda_1 = \cdots = \lambda_n = 0. \tag{A 3}$$

If this condition is not satisfied we may reduce the number of extensive variables so as to obtain linearly independent phase functions.

The functions $a^\mu(\zeta)$ being measurable, the integral

$$\Phi(\beta) \equiv \int_Z \exp\{-\beta_\mu a^\mu(\zeta)\}\, \mu(d\zeta) \tag{A 4}$$

is defined for arbitrary real parameters $(\beta_1, \ldots, \beta_n) \equiv \beta$ and

$$0 < \Phi(\beta) \leqslant +\infty. \tag{A 5}$$

As the integral of a convex function, $\Phi(\beta)$ is itself a convex function of the parameters β_μ. That is, if $0 \leqslant p \leqslant 1$ and $\beta 1, \beta 2 \in R^n$

$$\Phi\,[p\,\beta\,1+(1-p)\,\beta\,2]\leqslant p\,\Phi\,(\beta\,1)+(1-p)\,\Phi\,(\beta\,2). \tag{A 6}$$

It follows that every n-dimensional point set of the form

$$\{\beta\,|\,\Phi\,(\beta)<c\} \tag{A 7}$$

with c an arbitrary real number, if non empty, is convex.

We shall assume that the integral $\Phi\,(\beta)$ is finite if β belongs to the, necessarily convex, open subset D_β of R^n, and infinite otherwise. In the one dimensional case of the ordinary canonical distribution one usually assumes that the integral $\int \exp\{-\beta\,H\,(\zeta)\}\,\mu\,(d\,\zeta)$, $H\,(\zeta)$ – the Hamiltonian, is finite for all values of β in the open interval $0<\beta<\infty$, [Khinchin, 1949].

The function $a\,(\zeta)$ and the measure $\mu\,(d\,\zeta)$ induce a measure on the (Borel) subsets of the n-dimensional real space R^n: Let B be a (Borel) subset of R^n and B' the corresponding subset of Z, $\zeta \in B' \subset Z$ if and only if $a\,(\zeta)\in B \subset R^n$, and define the measure $V\,(\cdot)$ in R^n by

$$V\,(B)=\mu\,(B'). \tag{A 8}$$

Evidently $V\,(R^n) = \mu\,(Z) = +\infty$. The integral (A 4) over Z can be written as an integral over R^n:

$$\Phi\,(\beta)=\int\limits_{R^n} \exp\{-\beta_\mu\,a^\mu\}\,V\,(d\,a^1\,\ldots\,d\,a^n), \tag{A 9}$$

and the integral is convergent if $\beta \in D_\beta$.

The measure $V\,(I)$ of a finite interval is finite. This follows from the fact that the integral (A 9) is finite for some value of β. Thus the integral over I is finite, and the integrand has a positive lower bound on I. Hence $V\,(I)<\infty$. The assumption that the integral (A 9) is convergent for some value of β is thus a considerable restriction on the class of admitted phase functions $a\,(\zeta)$.[1]

The support F of the measure $V\,(d\,a)$ is defined as the minimal closed subset of R^n with the property $V\,(R^n-F)=0$. That is, F consists of the possible values of the function $a\,(\zeta)$. Hence any integral over R^n with respect to $V\,(d\,a)$ can be replaced by an integral over F.

The assumption that the n phase functions $a^1\,(\zeta), \ldots, a^n\,(\zeta)$ are linearly independent implies that F contains at least $(n+1)$ linearly independent points. Thus F is not contained in any linear subspace of R^n with dimension less than n. It follows that the convex hull F^* of F has the same property and thus F^* contains an open subset, or equivalently: the interior of the convex hull of the support of $V\,(d\,a)$, F^{*0}, is not empty.

For every value of the parameters β in the open region D_β, the function $\Phi\,(\beta)$ has derivatives of all orders. This can be shown in the same manner as it is demonstrated e.g. in [Widder, 1946] that the one-dimensional Laplace-

---

[1] In the quantum mechanical case the assumption would e.g. imply that the spectrum of the operators corresponding to A is discrete, and that each eigenvalue has only a finite degeneracy.

Stieltjes transform has derivatives of all orders in the interior of its region of convergence. With $-\beta_\mu$ in (A 9) replaced by the complex variable $i z_\mu$, the integral defines for $\operatorname{Im}\{z\} \in D_\beta$, $\operatorname{Re}\{z\}$ arbitrary, an analytic function $\Phi(-iz)$ of the n variables $z_1, \ldots, z_n$.

The Massieu-Planck function $Y(\beta)$ is defined by

$$-Y(\beta) = \log \Phi(\beta) \tag{A 10}$$

and hence from (A 5)

$$-\infty \leqslant Y(\beta) < \infty. \tag{A 11}$$

Like $\Phi(\beta)$ the function $Y(\beta)$ is convex.

For complex argument, $\operatorname{Im}(z) \in D_\beta$, we define $Y(-iz)$ by

$$\lim_{\substack{\operatorname{Re}\{z\}\to 0\\ \operatorname{Im}\{z\}=\beta}} Y(-iz) = Y(\beta), \quad -Y(-iz) = \log \Phi(-iz). \tag{A 12}$$

The generalized canonical probability distribution in R^n characterized by the parameter values β in D_β is defined by the measure

$$\exp\{Y(\beta) - \beta_\mu a^\mu\}\, V(d a^1 \ldots d a^n). \tag{A 13}$$

With $\operatorname{Im}\{z\} = \beta$, $\operatorname{Re}\{z\} = \xi$

$$\frac{\Phi(-iz)}{\Phi(\beta)} = \exp\{Y(\beta) - Y(-iz)\} = \exp\{Y(\beta) - Y(\beta - i\xi)\} \tag{A 14}$$

is the Fourier transform (characteristic function) of the probability measure (A 13). By the inversion theorem for the Fourier transform we have for an arbitrary bounded Borel subset B of R^n

$$\begin{aligned}\int_B \exp\{Y(\beta) - \beta_\mu a^\mu\}\, V(d a^1 \ldots d a^n) = \int_{R^n} [(2\pi)^{-n} \int_B \exp\{-i\xi_\mu a^\mu\} \times \\ \times d a^1 \ldots d a^n] \exp\{Y(\beta) - Y(\beta - i\xi)\}\, d\xi_1 \ldots d\xi_n.\end{aligned} \tag{A 15}$$

This is equivalent to

$$\begin{aligned}V(B) = \int_{R^n} [(2\pi)^{-n} \int_B \exp\{-i(\xi_\mu + i\beta_\mu) a^\mu\}\, d a^1 \ldots d a^n] \times \\ \times \exp\{-Y(\beta - i\xi)\}\, d\xi_1 \ldots d\xi_n,\end{aligned} \tag{A 16}$$

or $\quad V(B) = \displaystyle\int_{\operatorname{Re}\{z\} \in R^n} [(2\pi)^{-n} \int_B \exp\{-i z_\mu a^\mu - Y(-iz)\}\, d a^1 \ldots d a^n]\, d z_1 \ldots d z_n,$

which holds for every value of $\operatorname{Im}\{z\} = \beta$ in D_β.

The first order derivative of the Massieu-Planck function

$$\alpha^\mu(\beta) = \frac{\partial Y(\beta)}{\partial \beta_\mu} = \int_{R^n} \exp\{Y(\beta) - \beta_\nu a^\nu\}\, a^\mu\, V(d a^1 \dots d a^n), \tag{A 17}$$

is equal to the mean of the probability distribution (A 13).

The second order derivative of $Y(\beta)$

$$\alpha^{\mu\nu}(\beta) = \frac{\partial^2 Y(\beta)}{\partial \beta_\mu \partial \beta_\nu} = \int_{R^n} \exp\{Y(\beta) - \beta_\lambda a^\lambda\}\, (a^\mu - \alpha^\mu)\, (a^\nu - \alpha^\nu)\, V(d a^1 \dots d a^n) \tag{A 18}$$

is equal to the second central moment of the distribution (A 13). The corresponding real quadratic form is positive

$$\alpha^{\mu\nu}(\beta)\, \lambda_\mu\, \lambda_\nu \geqslant 0 \tag{A 19}$$

and definite (according to (A 2) and (A 3)):

$$\alpha^{\mu\nu}(\beta)\, \lambda_\mu\, \lambda_\nu = 0 \text{ implies } \lambda_1 = \dots = \lambda_n = 0, \tag{A 20}$$

otherwise we would have with non vanishing $(\lambda_1, \dots, \lambda_n)$

$$\lambda_\mu (a^\mu - \alpha^\mu) = 0$$

for every a in F with the possible exception of a subset with measure zero, contrary to the assumption of linear independence.

The "possible values" of the phase functions $a^\mu(\zeta)$ form the closed set F, the support of $V(\cdot)$. The mean values $\alpha^\mu(\beta)$ of the generalized canonical distribution (A 13) must for any $\beta \in D_\beta$ be contained in the convex hull, F^*, of F. We shall prove that conversely every value α in the interior of F^*, that is in F^{*0}, is the mean value of a uniquely determined generalized canonical distribution. A corresponding theorem has been proved by Khinchin, [Khinchin, 1949], for the case of the ordinary canonical distribution (one function $a^\mu(\zeta)$, bounded from below, and corresponding to the energy). A generalization of Khinchins theorem to several phase functions has been given by Grad, [Grad 1952]. Grad restricts himself to functions $a^\mu(\zeta)$ of the phase, $\zeta = (q_1 \dots q_k, p_1 \dots p_k)$ which induce a measure $V(\cdot)$ in R^n of the form $v(a)\, d a^1 \dots d a^n$, with a continuous density $v(a)$. His proof is incomplete.

The theorem which we shall prove may be given the following precise formulation:

If α^* is a point in the interior of the convex hull of the support of the measure $V(\cdot)$, $\alpha \in F^{*0}$, the n equations

$$\frac{\partial Y}{\partial \beta_\mu}(\beta) = \alpha^{*\mu} \text{ or } \frac{\partial}{\partial \beta_\mu}[Y - \alpha^{*\nu} \beta_\nu] = 0, \tag{A 21}$$

$Y(\beta)$ being defined by (A 10), (A 9), are satisfied for exactly one value of β contained in the convex, open set D_β.

For the proof we consider the centered phase functions $a^{\mu}(\zeta)-\alpha^{*\mu}$ which in the following we shall denote by $a^{\mu}(\zeta)$. Thus it is no restriction to assume that the origin $0=(0, \dots, 0)$ is an interior point of F^* and to consider instead of (A 21) the equations

$$\frac{\partial Y}{\partial \beta_{\mu}}(\beta)=0, \qquad \beta \in D_{\beta}. \tag{A 22}$$

This, on the other hand, is the condition for a stationary value of $Y(\beta)$ at the point β, which as a consequence of (A 19) must be a maximum. The theorem is proved if it can be shown that $Y(\beta)$ has a unique maximum or equivalently that $\Phi(\beta)=\exp\{-Y(\beta)\}$ has a unique minimum in D_{β}. Denoting by Φ_0 the infimum of $\Phi(\beta)$ in D_{β} we must show that there exists exactly one point β in D_{β} for which $\Phi(\beta)=\Phi_0$.

The n-dimensional β-point set

$$\{\beta \mid \Phi(\beta) \leqslant \Phi_0+\varepsilon\} \tag{A 23}$$

where $\varepsilon>0$ is arbitrary, is evidently contained in D_{β} and is non empty. As a consequence of (A 6) the set is convex and due to the fact that $\Phi(\beta)$ is continuous in D_{β} it is closed. If the set (A 23) is also bounded it follows that for some point β^* in the set $\Phi(\beta^*)=\Phi_0$.

In order to show that the set (A 23) is bounded we shall first prove that if β is an arbitrary fixed point in D_{β}, $\gamma \neq 0$ an arbitrary fixed point in R^n and t a positive variable, the continuous function $\Phi(\beta+t\gamma)$ of t increases infinitely with t. For sufficiently small t, $(\beta+t\gamma)\in D_{\beta}$, and $\Phi(\beta+t\gamma)<\infty$. If t increases there are two possibilities. Either $(\beta+t\gamma)$ reaches for some finite value of t the boundary of D_{β} or $\Phi(\beta+t\gamma)$ remains finite for all values of t. We shall see that also in the latter case the value of $\Phi(\beta+t\gamma)$ increases infinitely with t. The integral defining $\Phi(\beta+t\gamma)$ gives rise to the inequality

$$\begin{aligned}\Phi(\beta+t\gamma)&=\int_{R^n}\exp\{-t\gamma_{\mu}a^{\mu}\}\exp\{-\beta_{\mu}a^{\mu}\}\,V(da^1\dots da^n)\geqslant\\ &\geqslant \int_{-\gamma_{\mu}a^{\mu}>\delta>0}\exp\{-t\gamma_{\mu}a^{\mu}\}\exp\{-\beta_{\mu}a^{\mu}\}\,V(da^1\dots da^n)\geqslant\\ &\geqslant \exp\{t\delta\}\int_{-\gamma_{\mu}a^{\mu}>\delta}\exp\{-\beta_{\mu}a^{\mu}\}\,V(da^1\dots da^n).\end{aligned} \tag{A 24}$$

The inequality (A 24) holds for any $\delta>0$. To see that $\Phi(\beta+t\gamma)$ increases infinitely with t we have only to show that the last integral is non vanishing for some $\delta>0$. The integrand being positive, this is equivalent to $V\{a \mid -\gamma_{\mu}a^{\mu}>\delta\}>0$. According to our assumption $0=(0, \dots, 0)\in F^{*0}$. Now F^{*0} is an open set. Hence there exists a $\delta>0$ such that $F^{*0}\cap\{a \mid -\gamma_{\mu}a^{\mu}>\delta\}\neq\phi$. It follows that $F\cap\{a \mid -\gamma_{\mu}a^{\mu}>\delta\}\neq\phi$. But this implies, as the set $\{a \mid -\gamma_{\mu}a^{\mu}>\delta\}$ is open, according to the definition of F, that $V\{a \mid -\gamma_{\mu}a^{\mu}>\delta\}>0$, for otherwise we would have $F\cap\{a \mid -\gamma_{\mu}a^{\mu}>\delta\}=\phi$. Thus (A 24) it seen to imply that $\Phi(\beta+t\gamma)$ increases infinitely with t.

This is sufficient to assure that the set (A 23) is bounded, for taking an arbitrary fixed point β in the set its distance to the boundary in an arbitrary direction is a continuous function of the direction (i.e. a continuous function on the unit sphere). This distance function assumes its maximum value for some direction. Thus the set is contained in a sphere with finite radius and center in β.

Hence it has been established that the function assumes its minimum value in some point of D_β. Thus the equations (A 22) and consequently (A 21) can be satisfied. The uniqueness of the solution is a result of (A 20) and has been proved in paragraph 7.

We have found that for arbitrary $\alpha^* \in F^{*0}$ the function $Y(\beta) - \beta_\mu \alpha^{*\mu}$ has a unique maximum for $\beta = \beta^*$ where β^* is determined by

$$\frac{\partial Y}{\partial \beta_\mu}(\beta^*) = \alpha^{*\mu} \qquad (\mu = 1, \ldots, n). \tag{A 26}$$

Hence

$$Y(\beta^*) - \beta^*_\mu \alpha^{*\mu} \geqq Y(\beta) - \beta_\mu \alpha^{*\mu} \qquad (\beta \text{ arbitrary} \in D_\beta). \tag{A 27}$$

Introducing the entropy function $S(\alpha)$ (42), (43) we obtain

$$-S(\alpha^*) \geqq \frac{\partial S(\alpha)}{\partial \alpha^\mu}\alpha^\mu - S(\alpha) - \frac{\partial S(\alpha)}{\partial \alpha^\mu}\alpha^{*\mu},$$

or (A 28)

$$S(\alpha) - S(\alpha^*) \geqq \frac{\partial S(\alpha)}{\partial \alpha^\mu}(\alpha^\mu - \alpha^{*\mu}).$$

The last inequality (A 28) holds for arbitrary α and α^* in F^{*0}, so that by the same argument

$$S(\alpha^*) - S(\alpha) \geqq \frac{\partial S(\alpha^*)}{\partial \alpha^\mu}(\alpha^{*\mu} - \alpha^\mu), \tag{A 29}$$

and hence

$$\frac{\partial S(\alpha^*)}{\partial \alpha^\mu}[\alpha^\mu - \alpha^{*\mu}] \geqq S(\alpha^*) - S(\alpha) \geqq \frac{\partial S(\alpha)}{\partial \alpha^\mu}[\alpha^\mu - \alpha^{*\mu}]. \tag{A 30}$$

ACKNOWLEDGEMENT

I wish to thank professor Oskar Klein, head of this institute, for his kind interest in the work and constant encouragement.

Institute for theoretical physics, University of Stockholm, Stockholm.

REFERENCES

Cramér, H., 1928, On the composition of elementary errors, Skandinavisk Aktuarietidskrift p. 13 and p. 141.
——, 1937, Random variables and probability distributions, Cambridge University Press.
——, 1946, Mathematical methods of statistics, Princeton University Press.

EHRENFEST, P. and T., 1911, Begriffliche Grundlagen der statistischen Auffassung in der Mechanik, Encyklopädie der Mathematischen Wissenschaften IV 32.

EINSTEIN, A., 1910, Theorie der Opaleszenz von homogenen Flüssigkeiten und Flüssigkeitsgemischen in der Nähe des kritischen Zustandes, Annalen der Physik *33* 1275.

GIBBS, J. W., 1902. Elementary principles in statistical mechanics, Charles Scribner's Sons.

GNEDENKO, G. V. and KOLMOGOROV, A. N., 1954, Limit distributions for sums of independent random variables, Addison-Wesley.

GOURSAT, E., 1917, Cours d'analyse mathématique, Tome I, Gauthier-Villars.

GRAD, H., 1952, Statistical mechanics, thermodynamics, and fluid dynamics of systems with an arbitrary number of integrals, Comm. pure and applied mathematics *5* 455.

HALMOS, P. R., 1950, Measure theory, Van Nostrand.

KHINCHIN, A. I., 1949, The mathematical foundation of statistical mechanics, Dover.

MAXWELL, J. C., 1878, On Boltzmann's theorem on the average distribution of energy in a system of material points, Scientific papers *2* 713, Cambridge University Press.

MÜNSTER, A., 1953, Statistische Schwankungen und thermodynamische Stabilität, Zeitschrift f. Physik *136* 179.

——, 1959, Prinzipien der statistischen Mechanik, Handbuch der Physik III/2 p. 176.

PLANCK, M., 1932, Vorlesungen über Thermodynamik, Walter de Gruyter.

WIDDER, D. V., 1946, The Laplace transform, Princeton University Press.

Tryckt den 30 mars 1960

Uppsala 1960. Almqvist & Wiksells Boktryckeri AB